

---

---

# Storage Space Reporting

— Wei Yang —  
On behalf of the DDM team

---

---

# Storage protocols we are currently using

File access:

- file/posix : local access only
- xroot : local and remote
- HTTP : local and remote

File transfer:

- SRM + GridFTP : by far the most popular protocols
- xroot
- HTTP

Space report :

- SRM : mostly used. As we make SRM optional, **we need to find a new method for space reporting**
- GridFTP or HTTP : the new space reporting method need to work with either GridFTP or HTTP, or both

# ATLAS space reporting JSON

The DDM team proposed to use a JSON file to report space usage info

- It no longer ties space report to any protocol.
- It is a necessary step to make SRM optional
- Format:

```
"ATLASDATADISK":  
  {"status": "online"/"offline",  
   "status_message": "The report can not be created because ...",  
   "List_of_paths": ["/path1", "/path2"],    Note: this is a list!  
   "total_space": 5000000000,  
   "used_space": 2000000000,  
   "num_files": 123456,  
   "time_stamp": 1447936989}  
"ATLASSCRATCHDISK":  
{same}
```

- ATLAS\*DISK is no longer viewed as space token. They are simply names for storage areas/paths with specific purposes - for production data, for user data, etc.

# ATLAS space reporting JSON, cont'd

It is either a file in your storage that DDM will retrieve periodically, or generated on the fly (e.g. DPM 1.9 will provide this via http protocol)

Storage system providers (dCache, DPM, etc.) are also working on new mechanism to report space usage.

- They may or may not adopt ATLAS space reporting JSON.
- ATLAS will help their sites to use ATLAS space reporting json
  - <https://twiki.cern.ch/twiki/bin/viewauth/AtlasComputing/SRM2GridftpMigration>

Deployment status:

- US sites: AGLT2, BNL, LUCILLE, OU\_OCHEP, SLAC
  - AGLT2: <gsiftp://dcap.aglt2.org//pnfs/aglt2.org/atlasscratchdisk/space-usage.json>
- ES sites: PIC, IFAE

# More discussion with the DDM team

Additional work is needed to fully use the ATLAS space reporting JSON in RUCIO

- for example, total\_space

Special cases:

A large shared disk pool shared by all ATLAS\*DISK: **CEPH**, posix filesystem (**GPFS, Lustre**). They don't always trace "used space" and "num\_files", but

- RUCIO knows these numbers, can we use that?
- We generate monthly storage dump. Will that correct any inconsistency between what RUCIO knows and the actual site usage?
- If so, we don't need to modify the JSON schema, but how we interpret it.